

Three-Channel Convolutional Neural Network for Polarimetric SAR Images Classification

Wenqiang Hua¹, Member, IEEE, Wen Xie, Member, IEEE, and Xiaomin Jin²

Abstract—Terrain classification is an important topic in polarimetric synthetic aperture radar (PolSAR) image processing and interpretation. A novel PolSAR classification method based on the three-channel convolutional neural network (Tc-CNN) is proposed and this method can effectively take the advantage of unlabeled samples to improve the performance of classification with a small number of labeled samples. Several strategies are included in the proposed method. First, in order to take the advantage of unlabeled samples, a data enhancement method based on the neighborhood nearest neighbor propagation method is proposed to enlarge the number of labeled samples. Second, to increase the role of central pixel in convolutional neural network classification based on the pixel, a spatial weighted method is proposed to increase the weight of central pixel features and weak the weight of other types of pixel features. Third, a specific deep model for PolSAR image classification (named Tc-CNN) is proposed, which can obtain more scale and deep polarization information to improve the classification results. The experimental results show that the proposed method achieves a much better performance than the existing classification methods when the number of labeled samples is few.

Index Terms—Convolutional neural network (CNN), polarimetric synthetic aperture radar (PolSAR), terrain classification, three-channel convolutional neural network (Tc-CNN).

I. INTRODUCTION

IN RECENT years, polarimetric synthetic aperture radar (PolSAR) has caused a great deal of attention because it has a wide range of applications on earth remote sensing in many fields, such as marine environmental protection, target detection, and land cover classification [1]–[3]. PolSAR is high-resolution imaging equipment with multichannel, which records scattering echoes of different polarization modes with rich terrain information. It has the capabilities of all-weather and day-night. Over the past few decades, a large number of PolSAR systems (ESAR, TerraSAR-X, AIRSAR, ALOS2, and Radarsat-2) have been developed in and many of PolSAR images have been made publicly available.

Among the utilization of PolSAR data, land cover classification is much in demand. A number of PolSAR classification

Manuscript received April 16, 2020; revised June 25, 2020; accepted August 15, 2020. Date of publication August 20, 2020; date of current version September 4, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61901368, in part by the Scientific Research Plan Projects of Shaanxi Education Department [19JK0798], and in part by the Natural Science Foundation of Shaanxi Province under Grant 2019JQ-377 (Corresponding author: Wenqiang Hua.)

The authors are with the Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China (e-mail: huawenqiang@xupt.edu.cn; xiewen@xupt.edu.cn; xmjin@xupt.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3018161

methods, including target decomposition [4], [5], multiple statistical distributions [6], [7], and sparse representation [8], [9], have been developed in recent years. Generally, there are two main steps for PolSAR image classification: feature extraction and representation, and classifier designing and optimization.

For the first aspect, target decomposition methods based on the scattering mechanism are commonly used to extract PolSAR feature information, for instance, Pauli decomposition [4], H/a/A decomposition [5], Freeman three-component decomposition [6], Krogager decomposition [10], and Huynen decomposition [11]. The parameters of the polarimetric target decomposition are related to the physical properties of terrain object and have been utilized for unsupervised and supervised PolSAR image classification. Meanwhile, polarimetric features extracted by other model-based decompositions based on the basic scattering mechanism have also been explored for the PolSAR image classification. In addition, color features, texture features, and some statistic features have also been used to enhance the PolSAR image classification.

For the other hand, various classifiers based on machine learning, such as the complex Wishart classifier [12], Bayes classifier, support vector machine (SVM) classifier [13], decision tree [14], k -nearest neighbor (k -NN) classifier [15], and neural networks classifier [16], have been adopted for PolSAR image classification. What is more, the PolSAR image classification performance has been improved with the advances of classifier performance.

More recently, the emergence of deep learning techniques has opened up a new way for PolSAR image classification. The early deep learning method mainly includes three kinds of network structures: the convolutional neural network (CNN) [17], deep belief network (DBN) [18], and stacked sparse autoencoder (SSAE) [19]. At present, with the development of deep learning techniques, more and more deep learning methods have been applied for PolSAR image classification, such as clustering Wishart autoencoder [20], complex-valued deep fully convolutional network [21], and deep sparse tensor filtering network [22]. Compared with the traditional image classification methods, feature extraction and classifier design are two relatively independent processes. However, in the deep learning method, the process of feature extraction and classifier design is a whole. In the forward process, deep features are extracted by the deep network, and then the classifier is trained by deep features. In the process of backpropagation, according to the output of the classifier, the depth network is fine-tuned to make the extracted depth features more consistent with the

classification characteristics of the classifier and improve the classification effect.

However, the success of deep learning classification methods is subject to the accessibility of large amounts of labeled data. For PolSAR image, it is well known that the labeled samples are difficult to obtain in practice and require a great deal of manpower and material resources. Therefore, we seek to answer this critical question in this work: *How to achieve better performance with small samples when applying deep learning techniques in the PolSAR image classification.*

According to this problem, this article starts from two aspects to improve the classification accuracy of PolSAR images classification with a small number of labeled samples. One is to increase the utilization of spatial information. For the PolSAR image classification, spatial relation is very important information, which can effectively improve the classification performance of PolSAR [23]–[25]. CNN [17] as one of the main models of deep learning method, with its unique network structure, can effectively extract the deep spatial features. What is more, CNN has been widely used in image processing applications and achieved remarked application results [26]–[28]. In recently, the convolution network is used to process PolSAR data [29]–[31], which not only extract the deep features but also extract the spatial features of PolSAR data based on the convolution structure. However, the traditional CNN usually has only one channel; for image processing problems, the original features of images usually contain feature information of different sizes. The single convolutional kernel of single-channel CNN loses a part of the different scale features of the original image so that the extracted features are insufficient.

Therefore, in order to obtain the richer and more diverse spatial feature, a three-channel convolutional neural network (Tc-CNN) is proposed in this work. This model contains three independent CNN networks, and each convolution layer uses different convolution kernels to extract multiscale and diversity features. However, in the traditional PolSAR classification based on CNN, the input of CNN is the pixel block selected from the center of a pixel, and the output category is used to represent the category of the central pixel. Although this pixel block selection strategy takes into account the spatial relationship between the pixels, it weakens the influence of central pixels on their own categories. Accordingly, we proposed a spatial weighted (SW) method for the pixel block selection method, which increases the weight of the features of central pixels and weakens the weight of the features of different pixels in the block.

Another aspect is to increase the number of training samples through data enhancement (DE). According to the region consistency of the image, the pixels in adjacent regions have high similarity. Therefore, combined with the characteristics of the PolSAR image, we proposed a new DE method based on the neighborhood nearest neighbor propagation (N3P) method, which uses a small number of labeled samples to select the unlabeled samples in its neighborhood and enlarge the set of labeled samples.

Therefore, the main contributions of the present article are summarized as follows.

- 1) We proposed a new PolSAR image classification method with a small number of labeled samples. In particular, a Tc-CNN is designed to extract the deep polarimetric features and output class label predictions. This method is realized that deep learning method under a small sample can effectively solve the PolSAR image classification problem.
- 2) To reduce the influence of different types of pixels in the input pixel block on the central pixel, we propose an SW method for feature selection of pixel blocks to increase the influence of central pixel on their own classification and further improve the classification results.
- 3) In order to further improve the performance of the proposed method with the small number of labeled samples, a DE method based on N3P is employed to enlarge the set of training samples.
- 4) Finally, we apply the newly present approach to two real PolSAR datasets. The experimental results demonstrate that compared with the existing classification methods, the proposed method achieves state-of-the-art classification performance when the number of labeled samples is few.

The remainder of the article is organized as follows. Section II explains the proposed method. Section III shows the experimental design with real PolSAR data and the experimental results are provided in Section IV. This is followed by a discussion and conclusions for this article in Section V.

II. PROPOSED METHOD

A. Polarimetric Features

In general, each pixel in PolSAR data can be expressed in the form of a covariance or coherency matrix T as follows:

$$T = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix}. \quad (1)$$

Since T is a complex conjugate matrix and the diagonal elements are real numbers, we utilize the coherency matrix to define a new 9-D vector corresponding to this coherency matrix T as follows:

$$v1 = [T_{11}, T_{22}, T_{33}, Re(T_{12}), Im(T_{12}), Re(T_{13}), Im(T_{13}), Re(T_{23}), Im(T_{23})]. \quad (2)$$

Here, $Re(\cdot)$ and $Im(\cdot)$ represent the real and imaginary parts of a complex number, respectively.

However, in order to extract more abundant feature information with more differences, this article presents a three-channel convolution network for PolSAR classification. Therefore, the feature vector view is further divided into three small feature subsets (view1 = view(1 : 3), view2 = view(4 : 6), and view3 = view(7 : 9)), which are input to each channel. To illustrate the effectiveness of these three feature subsets, we analyzed the distribution of the view1, view2, and view3 to illustrate the effectiveness of three feature subsets, as shown in Fig. 1.

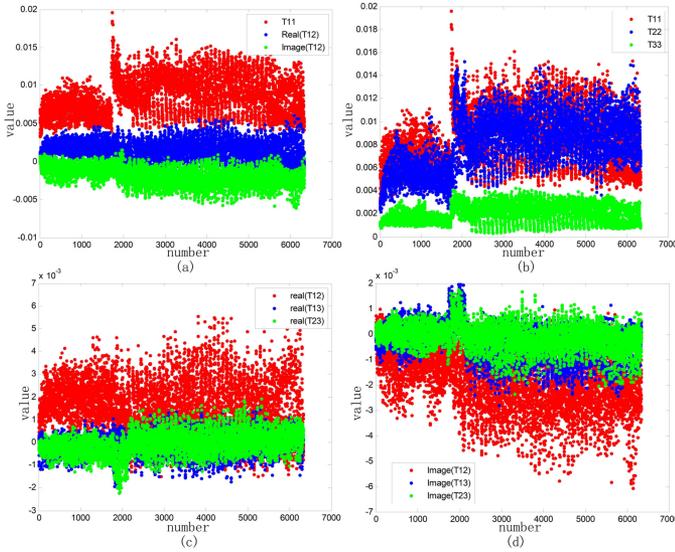


Fig. 1. Distribution of three feature subsets. (a) Difference between view1, view2, and view3. (b) Feature distribution of view1. (c) Feature distribution of view2. (d) Feature distribution of view3.

The original image of Flevoland areas is shown in Fig. 7(a). Fig. 7(b) shows the ground-truth map. In Fig. 1, we take terrain type of stem beans as an example to analyze the distribution of different elements of coherency matrix T . A total of 6000 pixels are selected from stem beans. From Fig. 1, we can clearly find that there are obvious differences between the diagonal elements, the real part, and the imaginary part of the complex number in the coherent matrix T . Therefore, the difference in the coherent matrix T may be of benefit to classification.

B. DE Based on N3P

In order to effectively utilize unlabeled samples to enhance the labeled sample set, the N3P method is adopted in the present article. The N3P method is based on the traditional k -NN algorithm.

For k -NN algorithm, one first needs to build a set of vertices and distance $G = (V, W)$, where each node in V corresponds to a data point, and W is the Euclidean distance between any two points, which represents the similarity between any two points. The smaller the value of W , the higher the similarity. For the PolSAR image, each pixel is considered a data point, and the set of vertices and edges G should be constructed by all pixels of the PolSAR image because PolSAR data obey the complex Wishart distribution. Therefore, the Wishart distance is used to express the similarity between any two points, and the expression is as follows:

$$W_{i,j} = \frac{1}{2} \{ \ln(|T_i|) + \ln(|T_j|) + \text{Tr}(T_i^{-1}T_j + T_j^{-1}T_i) \} \quad (3)$$

where T is the coherency matrix and $\text{Tr}()$ denotes the trace of a matrix. However, for PolSAR images, it takes a large amount of computation to calculate the distance between all pixels. Fortunately, Xi *et al.* [32] show that in the PolSAR image, the adjacent pixels in the same area are more likely to have

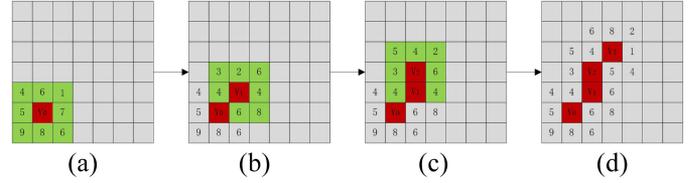


Fig. 2. Flowchart of the N3P method.

a similar label. Therefore, in order to reduce the computation and increase the reliability of selected vertices, we propose a novel N3P method, which is described as follows.

Step 1: Construct a set of vertices and distance G by each labeled pixel and its eight neighborhood pixels.

Step 2: Use labeled pixel as the vertices, the label is propagated to the nearest neighbor in its eight neighborhoods, and select the nearest neighbor to add to the vertex set V .

Step 3: Apply the new labeled pixel as the vertices to construct a new set of vertices and distance G .

Step 4: Repeat step 2 and step 3 until enough vertices are selected.

Fig. 2 shows the flowchart of the N3P method. The red point in Fig. 2(a) is the initial vertex V_0 , the green points represent the neighboring vertices of V_0 , and the gray points represent other non-neighborhood pixels. First, calculate the distance between the labeled pixel V_0 and its eight neighborhood pixels. The numeric value at each neighboring point denotes the Wishart distance between V_0 and that point, such as Fig. 2(a). Then, propagate the label to the nearest pixel, such as “1,” and add this pixel to the vertex set V . Take this pixel as a new labeled point V_1 , and calculate the distance between this point and its eight neighborhoods [see Fig. 2(b)]. Next, propagate label to point V_2 nearest to V_1 , as shown in Fig. 2(c). Repeat this process until enough vertices are labeled. In Fig. 2(d), the N3P method has propagated three times, and the number of labeled samples has expanded three times.

C. Spatial Weighted

For the PolSAR terrain classification, spatial information is very important information. Therefore, how to effectively utilize spatial information in the PolSAR image has been one of the research hot topics of PolSAR classification. CNN has made great achievements in the field of image processing by mining the correlation between data. However, in PolSAR data processing based on CNN, an image block is usually selected with a certain pixel as the center, and the final class of the center pixel of this region is determined with all the information of the whole region. However, if most of the pixels in this region and the central pixel belong to different categories, the final category may have errors with the real category. Therefore, this article studies the relationship between the center pixel and the neighborhood pixel in the selected image block to increase the weight of the center pixel and the similar pixel information in the neighborhood by weighting and weakens the weight of the pixel information in the neighborhood with a large difference. It not only considers the

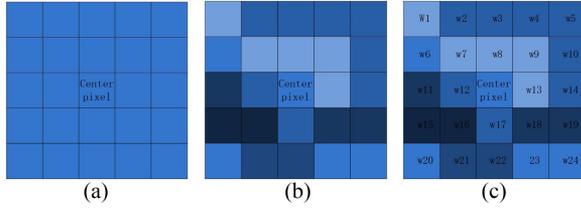


Fig. 3. Method of pixel block selection based on SW.

influence of the spatial pixel but also increases the contribution of the center pixel to its own category.

Fig. 3 shows the method of pixel block selection based on SW. The pixel block is selected with the center pixel, as the center determines the mark of the center pixel. Fig. 3(a) shows the pixel block selected under the ideal conditions: the central pixel and its neighborhood pixels belong to the same class. However, Fig. 3(b) shows the pixel block selected in the real PolSAR image. There are differences between the central pixel and its neighboring pixels, even different types. Therefore, it is very important to highlight the weights of the central pixel and its adjacent pixels and weaken the weights of different types of pixels in the neighborhood. Fig. 3(c) shows the pixel block selected after weighting, where the weight of the feature of each pixel in the pixel block is expressed as

$$SW_{ij} = \frac{w_{ii}}{w_{ij}} \quad (4)$$

$$w_{ij} = \frac{1}{2} \text{Tr}((V_i)^{-1} T_j) + \ln |V|. \quad (5)$$

w_{ij} denotes the similarity between the center pixel and its neighborhood pixel, and the smaller the value of w_{ij} , the higher the similarity. V_i denotes the coherency matrix of center pixels and T_j denotes the coherency matrix of neighborhood pixels. From (4), it can be seen that the more similar the neighborhood pixel is to the center pixel, the larger the value of w_{ii}/w_{ij} , the greater the weight, the greater the proportion of the pixel features, and the greater the impact on the final classification.

D. CNN Architecture

In this article, a Tc-CNN model is proposed to select diversified depth features and classification. In this model, each channel network is a single-channel CNN network, and each channel CNN network has different input features. Therefore, we will first introduce the CNN model in this section. As shown in Fig. 4, a CNN model generally consists of convolution layers, pooling layers, and fully connected layers.

Convolutional layers play an important role in feature extraction. In the deep CNN model, the initial convolutional layer captures the low-level features and the deeper convolution layers extract the abstract features. In the process of feature extraction, each unit of the convolutional layer is connected to a local patch in the feature maps of the previous layer by the convolutional kernels. In this process, all inputs in a feature map share the same convolutional kernels, and different feature maps use the different convolutional kernels. The convolution operation is as

follows:

$$x^l = g \left(\sum x^{l-1} * W^l + b^l \right) \quad (6)$$

where x^{l-1} denotes the input of the l th layer, W^l denotes the weight of the convolution kernel, b^l denotes the basis, $*$ represents the convolution operation, and $g(\cdot)$ denotes the non-linearity activation function, such as ReLU, tanh, sigmoid, and so on.

The pooling layer is also named subsampling layers. It aims to reduce the dimension of the previous layer feature map. There are mainly two kinds of pooling methods: average pooling and max pooling. Average pooling extracts the average value of the pooling region, and max pooling extracts the maximum value of the pooling region. Concrete operation is shown as follows:

$$x^l = g(\text{down}(x^{l-1}) + b^l) \quad (7)$$

where x^{l-1} is the input, and $\text{down}(x^{l-1})$ denotes pooling the features map of the layer $(l-1)$ th.

A fully connected layer is usually used at the end of the network. The last feature map is usually reshaped into a 1-D feature vector. Then, the 1-D feature vector is taken as the input of the output layer and output the final results.

E. Architecture of the Proposed Tc-CNN Model

This section describes the proposed method in detail. Fig. 5 shows the framework of the complete process of the proposed method. As can be seen from Fig. 5, the proposed method consists of two processes: training and classification. In the training process, there exist three steps. First, preprocessing the initial polarimetric data. This process contains two parts: DE and SW. Second, the input polarimetric features are divided into three feature subsets based on the statistical distribution of features. These three data subsets are used to separately extract different deep polarimetric features. Finally, a three-channel convolutional network model is used to extract multiscale depth features. In the classification process, the trained Tc-CNN classifier is used to analyze all unlabeled samples and determine their final labels. The detailed process is described as follows.

- 1) Randomly select the initial training sample set L .
- 2) The proposed DE method is used to strengthen the initial sample set L and then obtain the enhanced training sample set L' .
- 3) Each pixel in the PolSAR image is taken as the center, and the block of pixels of size $n \times n$ is selected as the basic input unit for the network model.
- 4) Each pixel block is processed by an SW method to increase the impact of the central pixel on the final results.
- 5) Three CNN networks are used to construct a three-channel convolutional network model. In this three-channel network, each CNN network uses different input information (view1, view2, and view3), and the number of convolutional layers and convolution kernels is different in each CNN network.
- 6) The inception model is used to contact the feature maps of the output of three different CNN networks. Then, these

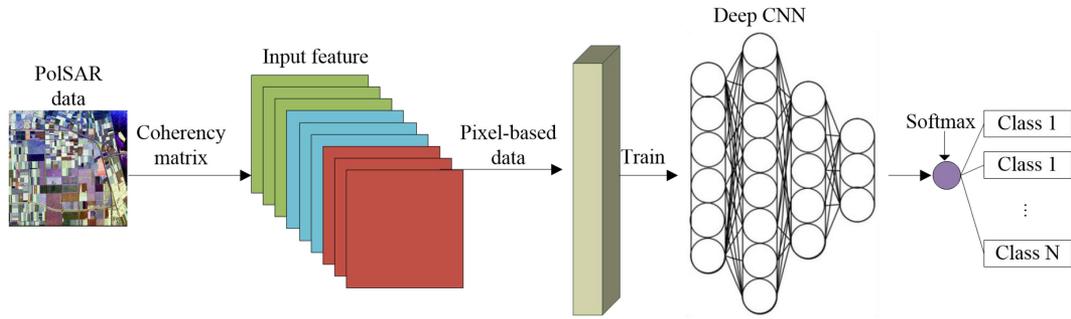


Fig. 4. CNN model.

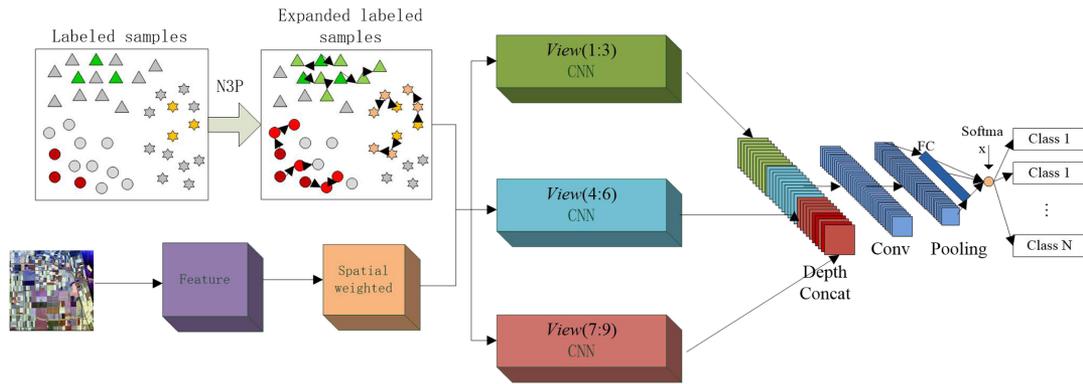


Fig. 5. Framework of the proposed method.

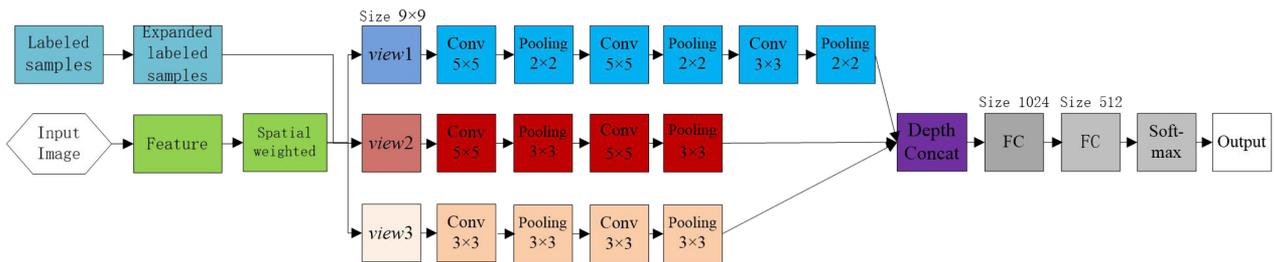


Fig. 6. Structure and parameters of the proposed method.

contacted features map as the input to the fully connected layer and output the final results.

- 7) Input the training sample and train the proposed Tc-CNN network.
- 8) Use the trained Tc-CNN network to classify the whole PolSAR image.

III. EXPERIMENTAL DESIGN

In the present work, a number of experiments are carried out with the Tensorflow2.0 on a desktop computer that runs Windows 10 operating system with an Intel(R) Core(TM) CPU processor (3.20 GHz) and 16 GB memory. The structure and some parameters of the proposed method are shown in Fig. 6. The size of the input pixel block is set to 9×9 , the learning

rate is set to 0.1, weight decay is set to 0.002, the pooling mode is mean pooling, and ReLU is used as a nonlinearity activation function. Two real PolSAR datasets are used as shown in the following text. By comparing the performance of different classification methods, the overall accuracy (OA) and the kappa statistics of classification results are adopted to demonstrate the effectiveness of the proposed classification method.

A. Datasets

1) *AIRSAR Data in Flevoland*: The first dataset is an L-band PolSAR data acquired by the NASA/JPL ARISAR system in Flevoland, The Netherlands, in August 1989 [28]. Fig. 7 shows both the Pauli image and the ground-truth image of this data.

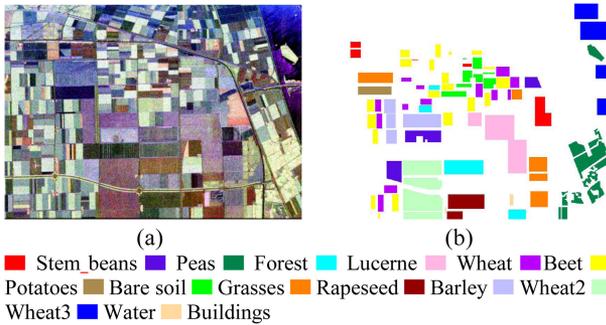


Fig. 7. AIRSAR dataset in Flevoland I. (a) Pauli image. (b) Ground-truth image.

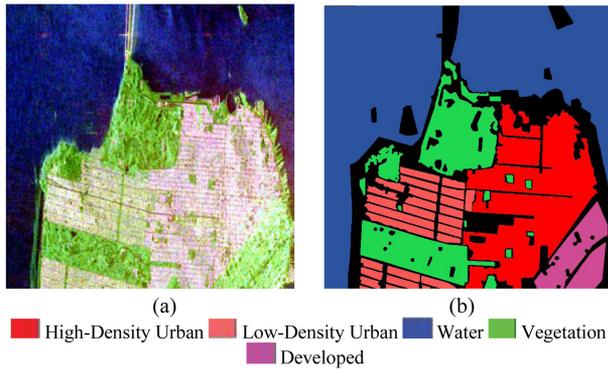


Fig. 8. RADARSAT-2 data in San Francisco. (a) Pauli image. (b) Ground-truth image.

The ground-truth image is obtained from the article presented in [33]. The size of the Pauli image data is 750×1024 pixels, while the size of the ground-truth data is around 167 712 pixels. The pixel size of the Pauli image is 6.6 m in the slant range direction and 12.1 m in the azimuth direction. There are 15 various terrain types marked in the ground truth, and most of which are agricultural classes, including water, rapeseed, grasses, bare soil, potatoes, beet, wheat, Lucerne, forest, peas, buildings, and stem beans.

2) *RADARSAT-2 Data in San Francisco*: Fig. 8 shows the Pauli image and the ground-truth image of the San Francisco dataset. This dataset is a C-band PolSAR data acquired by the RADARSAT-2 system in San Francisco, CA, USA, in April 2008. The ground-truth image is obtained from the article presented in [34]. The size of the Pauli image data is 1300×1300 pixels, while the size of the ground-truth data is around 1 341 917 pixels. The pixel size of the Pauli image is 10 m in the slant range direction and 5 m in the azimuth direction. As marked in the ground-truth image [see Fig. 8(b)], there are five different terrain types, including water, man-made, vegetation, low density, and high density.

B. Key Parameters in the Training Process

Two key parameters, including the number of propagation in the N3P method and the labeled samples in the training process, dominate the performance of the proposed method. In this

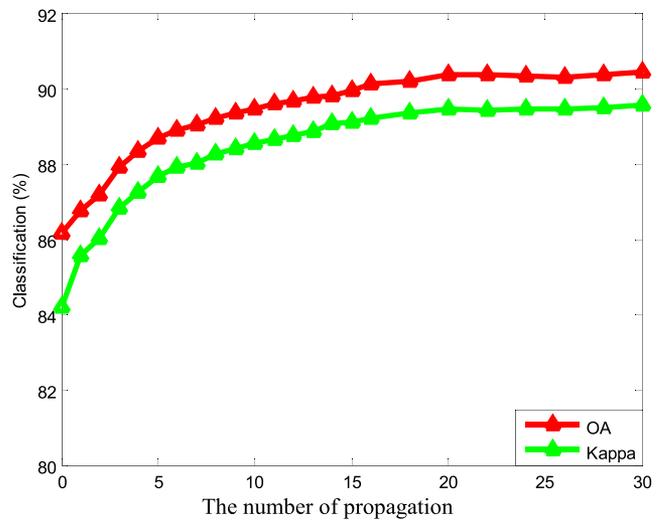


Fig. 9. Effects of the number of propagation on the OA and kappa statistics.

section, the effects of those two parameters on the classification results of the proposed method are investigated experimentally with the Flevoland dataset.

1) *Number of Propagation in the N3P*: In this section, we only consider the DE method and do not consider the effect of SW. Fig. 9 shows the OA and the kappa statistics of the classification results of the Flevoland dataset given by the proposed method as a function of the number of propagation used in the N3P method. In Fig. 9, the left data point (the number of propagation is equal to zero) of each curve corresponds to the case that the N3P method is not used in the training process. The OA and kappa statistics at this point have the lowest values for each curve. The data points with the number of propagation greater than zero refer to the cases where the proposed propagation method is applied and 1% labeled samples are used in the initial training dataset. As can be seen from Fig. 9, both the OA and kappa statistics increase as the number of propagation increases. This indicates that the proposed N3P method is an effective DE method, which effectively increases the number of training samples and improves the classification results.

In addition, Fig. 9 shows that the OA and kappa statistics of the classification results increase rapidly with increasing the number of propagation in a nonlinear fashion until the number of propagation reaches 10, and OA increased by 3.31%. However, as the number of propagation is greater than 10, both curves of the OA and kappa statistics become nearly flat. OA increased by 0.74% when the number of propagation increases from 10 to 20, and OA increased by 0.24% when the number of propagation increases from 20 to 30. This is mainly because the similarity between the root nodes and their subnodes might be weakened when the number of propagation increases. This weakening makes the enlarged labeled samples not play the role of expanding the labeled sample set. Thus, it does not obviously improve the classification performance of the network. In addition, with the increase of the number of propagation, the classification accuracy increases very little when the number of propagation increases from 10 to 30, but because of the increase of the

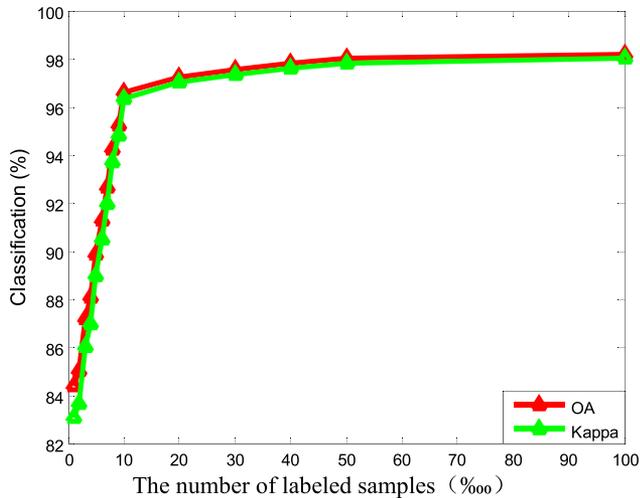


Fig. 10. OA and kappa statistics of the classification results of three PolSAR datasets vary with the number of labeled samples used in the proposed method.

number of propagation, the number of training samples also increases rapidly, which makes the time cost of training network increasing rapidly. Therefore, considering both classification accuracy and time cost, the number of propagation between 10 to 12 is suggested in the proposed method.

2) *Number of Labeled Samples*: Fig. 10 shows the OA and kappa statistics of the classification results of the Flevoland dataset as a function of the number of labeled samples used in the proposed method. In this section, the number of propagation is set to 10. The classification results are shown in Fig. 10. As can be seen from Fig. 10, the OA and kappa statistics are 84.41% and 0.8308, respectively, when 1 labeled samples (the left point in the figure) are used for the dataset of the Flevoland. The values of the OA and kappa statistics become 96.63% and 0.9633, respectively, as the number of labeled samples is 1‰. When the number of labeled samples is 1‰, the OA is 98.19%, and the kappa statistics is 0.9803. These results show that the OA increases 12.22% as the number of labeled samples used increases from 1 to 1‰ but only increases 1.56% as the number of labeled samples used increases from 1‰ to 1%. This is mainly because the fewer the labeled samples used in the training process, the greater the role of the proposed method, resulting in greater changes in the classification results. Thus, the proposed method is very effective for the PolSAR data where the labeled samples are few. It also indicates that with the increase of labeled samples used, the advantage of the proposed method is gradually weakened. To increase the efficiency of the proposed method and obtain satisfactory classification results, we suggest that 1‰ number of the labeled samples are used of the Flevoland dataset.

C. Effects of Speckle Filtering

To investigate the effect of suppressing speckle noise on classification results, Lee filter [35] with a 7×7 window is used. In this section, the Flevoland dataset is used as an example, 1‰ labeled samples are used as the training sample, and the number of propagation is set to 10. Fig. 11 and Table I presents the

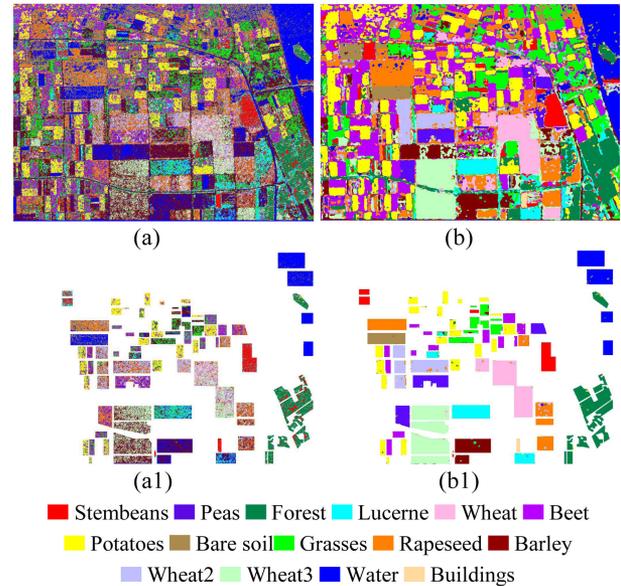


Fig. 11. Classification results of the Flevoland dataset. (a) Proposed method without a filter. (b) Proposed method with a filter. (a1) and (b1) are masked results according to the ground truth of (a) and (b), respectively.

TABLE I
CLASSIFICATION ACCURACY (%) OF THE FLEVLAND DATASET WITH AND WITHOUT FILTER

Method region	With Filter	Without Filter
Stem beans	99.12	73.59
Rapeseed	94.37	36.51
Bare soil	99.57	25.25
Potatoes	98.25	68.14
Beet	97.13	50.42
Wheat 2	93.12	47.30
Peas	97.72	40.96
Wheat 3	99.06	47.78
Lucerne	96.13	43.36
Barley	97.80	50.23
Wheat	98.44	24.44
Grasses	92.79	35.62
Forest	98.84	71.58
Water	98.88	96.09
Building	88.30	5.44
OA	96.63	47.78
Kappa	0.9633	0.4354

Bold number indicates the best result in each row.

classification results of the proposed method with and without the filter method.

As can be seen from Fig. 11 and Table I, the classification results of the proposed method without a filter are very poor. Its OA is only 47.78%, which is 48.85% lower than the proposed method with filter. This poor performance is mainly caused by two aspects. On the one hand, due to the influence of speckle noise, pseudolabels are added in the N3P process. This causes

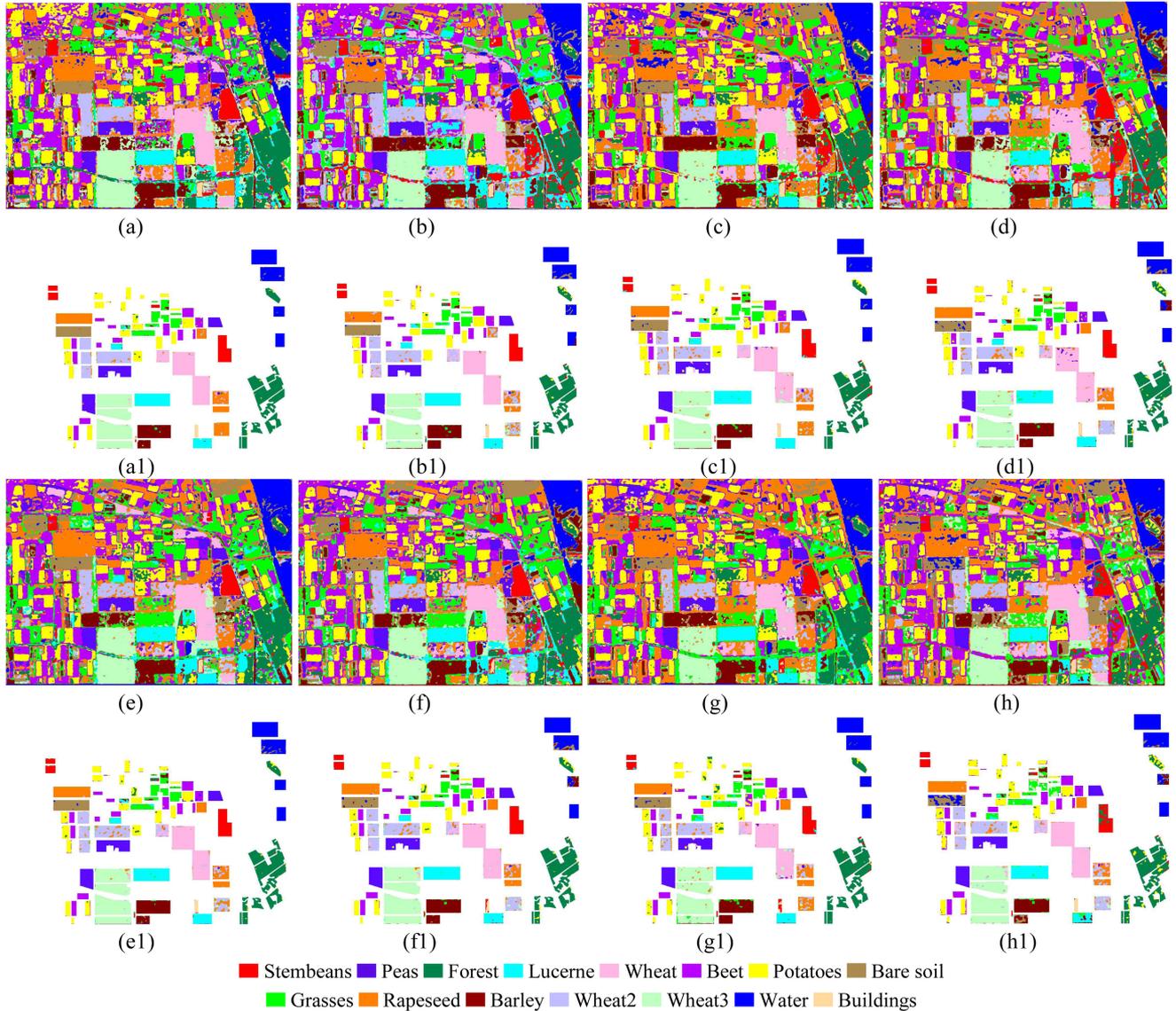


Fig. 12. Classification results of the Flevoland dataset. (a) Proposed method. (b) Tc-CNN+DE. (c) Tc-CNN+SW. (d) Tc-CNN. (e) CNN+DE+SW. (f) CNN+DE. (g) CNN+SW. (h) CNN. (a1), (b1), (c1), (d1), (e1), (f1), (g1), and (h1) are masked results according to the ground truth of (a), (b), (c), (d), (e), (f), (g), and (h), respectively.

that the newly added training samples contain an incorrect label, and this incorrect label will reduce the effectiveness of the proposed method, especially when the labeled samples are few. On the other hand, because there is no effective suppression of the speckle noise, it makes the training of neural networks more difficult, resulting in poor performance. Therefore, it indicates that the proposed method is very sensitive to speckle noise.

IV. EXPERIMENTAL RESULTS

In this section, the classification results of the two above-mentioned real PolSAR datasets given by the proposed method are presented and verified by the ground-truth measurements of those PolSAR datasets. The experiments are repeated ten times and the average values are used as the final classification results.

The experimental results are also compared with that obtained with a number of existing classification methods.

A. Classification Results of the Flevoland Acquired by AIRSAR in 1989

Fig. 12 shows the classification maps of the Flevoland dataset with different classification methods. Fig. 12(a) shows the results of the proposed method. Fig. 12(b) shows the classification of the method of the Tc-CNN plus only the DE (named Tc-CNN+DE). Fig. 12(c) shows the classification of the method of the Tc-CNN plus only the SW (named Tc-CNN+SW). Fig. 12(d) shows the classification of the method of the Tc-CNN (named Tc-CNN). Fig. 12(h) shows the classification of the traditional CNN [17] method. Fig. 12(e) shows the classification of CNN plus the

TABLE II
CLASSIFICATION ACCURACY (%) OF THE FLEVLAND AREA ACQUIRED BY AIRSAR

Method region	Proposed Method	Tc-CNN +DE	Tc-CNN +SW	Tc-CNN	CNN+DE +SW	CNN+DE	CNN+SW	CNN
Stem	99.12	95.36	94.60	96.61	93.53	97.63	93.70	83.81
Rapeseed	94.37	66.76	84.17	74.88	79.77	73.92	85.37	67.52
Bare soil	99.57	97.85	96.16	89.98	98.34	94.77	93.70	60.72
Potatoes	98.25	97.12	97.63	96.61	93.74	96.30	92.93	90.75
Beet	97.13	98.33	89.86	92.62	95.11	98.35	91.97	95.25
Wheat 2	93.12	88.59	86.07	85.31	87.60	88.56	83.20	86.34
Peas	97.72	97.84	96.24	99.13	99.12	98.81	96.57	98.46
Wheat 3	99.06	97.59	97.50	96.28	97.24	98.12	94.99	97.01
Lucerne	96.13	97.74	95.70	93.74	96.16	94.96	91.05	80.54
Barley	97.80	96.22	94.67	97.12	96.58	96.60	92.01	85.62
Wheat	98.44	91.82	92.02	88.88	90.31	91.76	89.89	91.36
Grasses	92.79	89.46	89.76	86.54	88.42	89.90	80.93	64.86
Forest	98.84	95.17	93.90	94.23	96.53	94.98	96.80	87.49
Water	98.88	93.02	98.63	88.34	96.21	86.12	99.70	91.82
Building	88.30	85.31	88.71	82.86	81.77	52.93	28.03	85.44
OA	96.63	92.54	93.04	90.88	92.70	90.25	87.39	84.47
Kappa	0.9633	0.9188	0.9242	0.9007	0.9205	0.8938	0.8626	0.8307

Bold number indicates the best result in each row.

DE and SW (named CNN+DE+SW). Fig. 12(f) shows the classification of CNN plus only the DE (named CNN+DE). Fig. 12(g) shows the classification of CNN plus only the SW (named CNN+SW). Table II presents the classification accuracy of each category with different classification methods and the OA for the Flevoland dataset. The results shown herein are obtained with 1% randomly selected labeled samples, and the number of propagation in the N3P is set to 10.

As can be seen from Table II, the proposed method performs the best in the majority of categories. The classification accuracy of 11 categories in the proposed method is higher than 95%, and the lowest classification accuracy of one category is 88.30%. In contrast, only five categories in the Tc-CNN method can achieve the classification accuracies above 95%, while the lowest classification accuracy of one category is only 74.88%. In the traditional CNN method, the classification accuracies of three categories are higher than 95%, and the classification accuracies of three categories are lower than 70%. In the CNN+DE+SW method, the classification accuracy of eight categories is higher than 95%, while the lowest classification accuracy of one category is 79.77%. Therefore, by comparing the proposed method with the traditional CNN method, one can see that the classification accuracy of the proposed method is 12.16% higher than the traditional CNN method. This indicates that the proposed method is an effective classification method that can significantly improve the classification accuracy of PolSAR data. Compared the Tc-CNN method with the CNN method, it is obvious to find that the classification accuracy of the Tc-CNN method is 6.41% higher than the CNN method. This indicates that the proposed Tc-CNN network is better than the CNN network for the PolSAR classification. This is mainly because the Tc-CNN network obtains the richer and more convolution features by different channels. These richer and diverse features improve the classification results of PolSAR data. Compared

the Tc-CNN+DE method, Tc-CNN+SW method, and TC-CNN method, one can see that the classification accuracy of the Tc-CNN+DE method is 1.66% higher than the Tc-CNN method, and the Tc-CNN+SW method is 2.16% higher than the Tc-CNN method. This indicates that both the proposed DE method and the SW method can significantly improve the classification results of the Tc-CNN method. Comparing the CNN+DE+SW method, CNN+DE method, CNN+SW method, and CNN, the classification accuracy of CNN method is 8.23%, 5.78%, and 2.92% lower than CNN+DE+SW, CNN+DE, and CNN+SW, respectively. This indicates that the proposed DE method and the SW method are not only effective for the proposed method but also effective for the traditional CNN method.

By comparing the classification accuracy of each category, as shown in Fig. 12, one can see that the classification accuracies of the proposed method in the stem, rapeseed, bare soil, potatoes, peas, wheat 2, wheat 3, Lucerne, barley, wheat, grasses, and forest are obviously higher than the other classification methods. For kappa statistics, the proposed method is 0.9633, which is notably higher than the other comparison methods. Therefore, we can also conclude that the proposed method is an effective method when the number of labeled samples is few. What is more, the proposed DE method and the SW method are both effective methods and can effectively improve the classification accuracy, especially when there are few labeled samples.

B. Classification Results of the San Francisco Acquired by Radarsat-2 in 2008

In this experiment, 1 labeled samples are randomly selected as the training samples. Fig. 13(a)–(h) shows the classification maps of different methods. The classification accuracy of these methods is given in Table III.

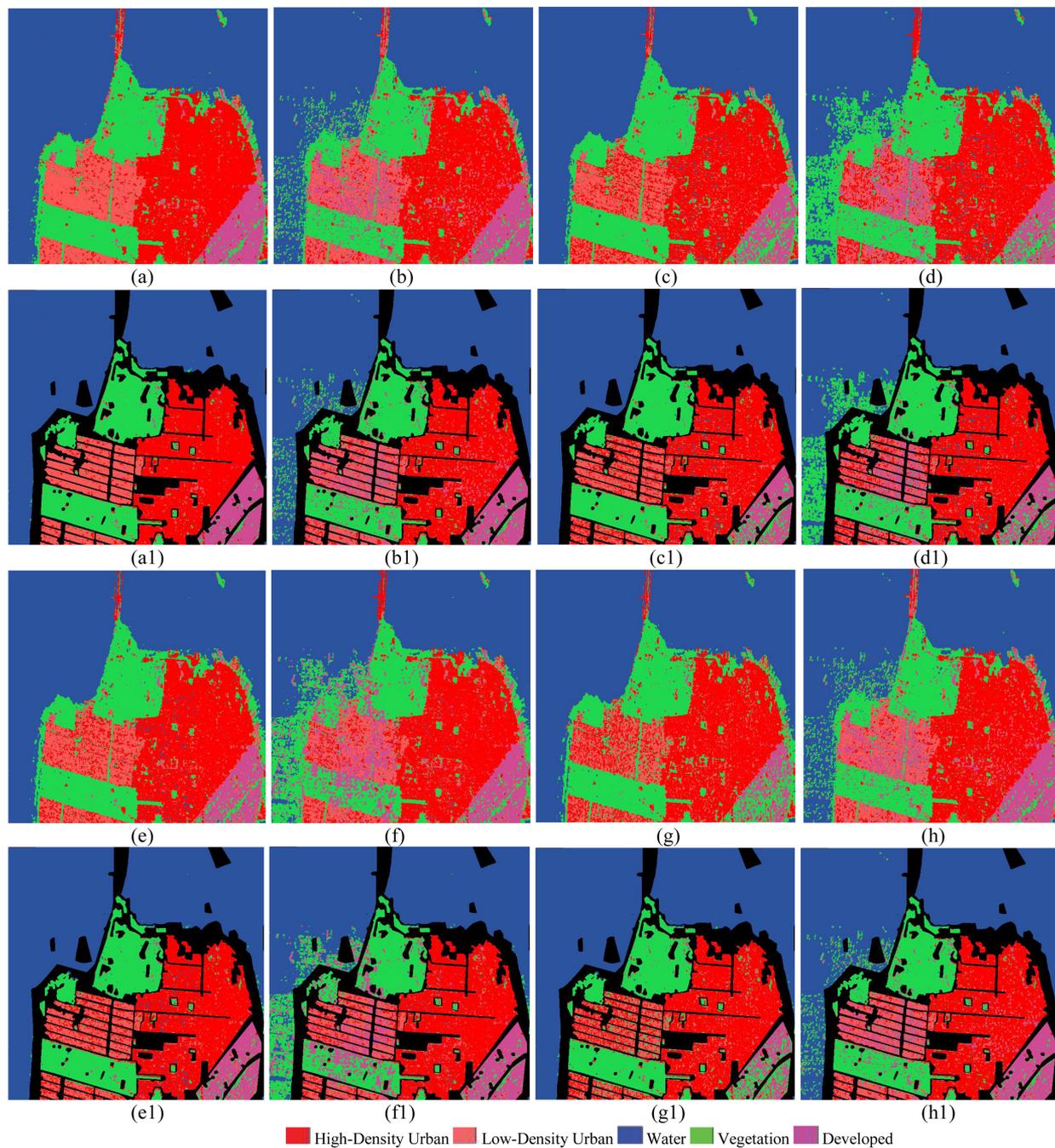


Fig. 13. Classification result of the San Francisco data acquired by Radarsat-2. (a) Proposed method. (b) Tc-CNN+DE. (c) Tc-CNN+SW. (d) Tc-CNN. (e) CNN+DE+SW. (f) CNN+DE. (g) CNN+SW. (h) CNN. (a1), (b1), (c1), (d1), (e1), (f1), (g1), and (h1) are masked results according to the ground truth of (a), (b), (c), (d), (e), (f), (g), and (h), respectively.

Table III presents that the overall classification accuracy of the proposed method is best than the other comparison methods. From Fig. 13 and Table III, it is clearly found that the classification accuracy of the proposed method is 93.64%, which is 9.78%, 10.51%, and 17.6% higher than Tc-CNN+DE, Tc-CNN+SW, and Tc-CNN methods, respectively. This indicates that each part of the proposed method can effectively improve the classification accuracy of the Tc-CNN network for PolSAR data. From Fig. 13 and Table III, it is clearly found that the

classification accuracy of the Tc-CNN method is only 44.80% in the low-density area and 69.17% in the developed area. However, the classification accuracy of the proposed method is 88.54% in the low-density area and 90.90% in the developed area, which is 43.47% and 21.73% higher than the Tc-CNN method, respectively. This further shows that a large number of unlabeled samples and spatial information play an important role in the PolSAR classification method and the effectiveness of the proposed method. Compared CNN+DE with CNN, the

TABLE III
CLASSIFICATION ACCURACY (%) OF THE SAN FRANCISCO AREA ACQUIRED BY RADARSAT-2

Method region	Proposed Method	Tc-CNN +DE	Tc-CNN +SW	Tc-CNN	CNN+DE +SW	CNN+DE	CNN+SW	CNN
Water	99.88	95.68	99.99	88.95	99.78	88.07	99.87	98.29
Vegetation	94.32	87.96	94.77	89.12	93.35	75.62	92.90	79.01
Low-Density	88.54	60.87	64.81	44.80	70.19	58.57	54.69	69.27
High-Density	94.56	87.28	87.06	88.18	90.89	89.65	87.32	62.52
Developed	90.90	87.52	69.01	69.17	77.08	83.30	50.03	59.63
OA	93.64	83.86	83.13	76.04	86.26	79.04	76.96	73.74
Kappa	90.40	76.03	74.40	64.83	79.14	69.57	65.01	60.83

Bold number indicates the best result in each row.

classification accuracy of the CNN+DE method is 5.3% higher than that of the CNN method after combined with individual DE. Comparing CNN+SW with CNN, the classification accuracy is 3.22% higher than that of CNN method after combined with individual SW. This conclusion also proves the validity of the proposed DE method and the SW method.

By comparing the kappa statistics, the kappa statistics of the proposed method is 0.9040, which is higher than the other comparison methods. This shows that the regional consistency of this method is obviously better than other comparison methods. What is more, by comparing Fig. 13, the regional consistency of the proposed method is better than other methods in the water area, vegetation area, low-density area, and developed area.

C. Comparisons With Existing Methods in Terms of the Number of Labeled Samples

In this part, the first dataset Flevoland is used to verify the effectiveness of the proposed method and compare the proposed method with the existing five traditional classification methods: CNN [17], SVM [13], sparse Autoencoder (SAE) [19], DBN [18], and Wishart classifier [12]. In this section, we select different proportions of labeled samples (N_l) to test the proposed method and other comparison methods. Fig. 14 shows the classification accuracy of different methods when different proportions of labeled samples are used as training samples. In order to better show the effect of Fig. 14, the abscissa is expressed in logarithmic form. In Fig. 14, the color notation (N_l, OA) on the curves represents the coordinates of the corresponding points, where N_l denotes the proportions of labeled samples, and OA denotes the classification accuracy. (N_l, OA) represents the classification accuracy of the points when selecting certain labeled samples as training samples. From Fig. 14, it can be clearly seen that the classification accuracy is higher than 90%, the proposed method needs 5 ($5 \approx 10^{0.6990}$) labeled samples, the traditional CNN method needs 40 ($40 \approx 10^{1.6021}$) labeled samples, the SAE method needs 80 ($80 \approx 10^{1.9031}$) labeled samples, and the SVM method needs 1000 ($1000 \approx 10^3$) labeled samples. However, when the proportion of training samples is 1000, the classification accuracy of DBN and Wishart classifier is 83.39% and 80.22%, respectively. Moreover, when the number of labeled samples is very few about 1, the classification accuracy of the proposed method is 84.04%, while

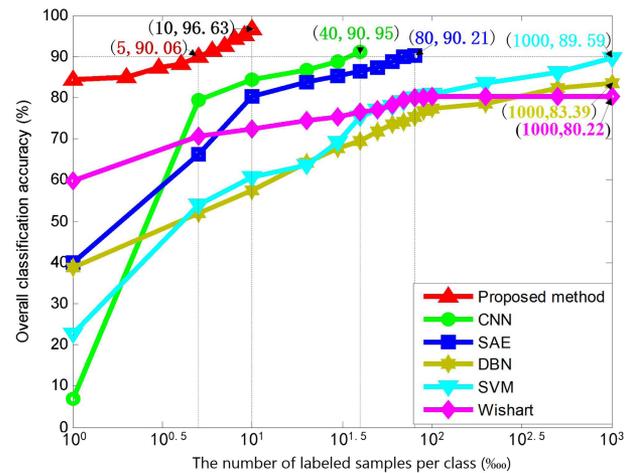


Fig. 14. Classification result of the Flevoland with different proportions of labeled samples. (In (N_l, OA) , N_l denotes the proportions of labeled samples, and OA denotes the classification accuracy.)

the classification accuracy of other methods is less than 60%. This is mainly because CNN, SVM, SAE, DBN, and Wishart methods require sufficient labeled samples as training samples. When the labeled samples are few, the classification accuracy of these five methods is poor. When the labeled samples are not sufficient, increasing the number of labeled samples means the classification accuracy of these five methods is also increased. This also indicates that when the labeled samples are few, the proposed method can obtain better classification results than the other comparison methods, but with the increasing of labeled samples, the classification accuracy of the traditional methods will gradually increase and finally approach the classification accuracy of the proposed method.

V. CONCLUSION

This article has presented an effective deep network model for the PolSAR image when labeled samples are few, which combines the traditional CNN network and spatial information of PolSAR data. In the proposed method, we proposed a DE method based on N3P to enlarge the labeled samples. Then, in order to strengthen the influence of the center pixel in the selected pixel block on its final category, an SW method is proposed to

weight the features of different pixels in a pixel block to weaken the feature weights of pixels that are weakly similar to the central pixel. Finally, we proposed a new three-channel network model to obtain more scale and deep polarization information to improve the classification results.

The experiments with two real PolSAR datasets showed that the proposed method is able to achieve satisfactory classification accuracy, especially when training samples are few. The experimental results show that the proposed method is an effective method and superior to other contrasting methods. In these two datasets, not only the overall classification accuracy of the proposed method is higher than the other contrasting methods but also the accuracy of the proposed method performs best in most classes. Furthermore, the experiments with the different training samples showed that the proposed method is able to achieve better classification results than the traditional methods. In addition, we analyzed the influence of the DE and SW on the results of the experiment. Fig. 9 indicates that the proposed DE method is reliable, and the performance of the classifier is gradually improved by the number of propagation to expand the label samples set. Figs. 12 and 13 indicate that the spatial information can effectively improve the classification results of PolSAR. In our future work, we will focus on the deep learning method of semisupervised learning to solve the PolSAR classification problem in small samples.

REFERENCES

- [1] F. S. Pramudya, J. Pan, and A. T. Devlin, "Estimation of significant wave height of near-range traveling ocean waves using Sentinel-1 SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1067–1075, Apr. 2019.
- [2] C. O. Dumitru, G. Schwarz, and M. Datcu, "SAR image land cover datasets for classification benchmarking of temporal changes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1571–1592, May 2018.
- [3] G. Liu, X. Zhang, and J. Meng, "A small ship target detection method based on polarimetric SAR," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 2938.
- [4] S. R. Cloude and E. Potter, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 498–518, Mar. 1996.
- [5] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, 1997.
- [6] A. Freeman and S. L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 963–973, May 1998.
- [7] J. J. Yin, X. Liu, J. Yang, C.-Y. Chu, and Y.-L. Chang, "PolSAR image classification based on statistical distribution and MRF," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 1027.
- [8] B. Ren, B. Hou, J. Zhao, and L. Jiao, "Sparse subspace clustering-based feature extraction for PolSAR imagery classification," *Remote Sens.*, vol. 10, no. 3, 2018, Art. no. 391.
- [9] N. Zhong, W. Yang, A. Cherian, X. Yang, G.-S. Xia, and M. Liao, "Unsupervised classification of polarimetric SAR images via Riemannian sparse coding," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5381–5390, Sep. 2017.
- [10] E. Krogager, "New decomposition of the radar target scattering matrix," *Electron. Lett.*, vol. 26, no. 18, pp. 1525–1527, Aug. 1990.
- [11] J. R. Huynen, "Phenomenological theory of radar targets," Ph.D. dissertation, Dept. Elect. Eng., Math Comput. Sci., Univ., Technolo., Delft, The Netherlands, 1970.
- [12] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D.-L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, Sep. 1999.
- [13] Y. Long and X. Liu, "SVM lithological classification of PolSAR image in Yushigou Area, Qilian," *Sci. J. Earth Sci.*, vol. 3, no. 4, pp. 128–132, 2013.
- [14] T. M. Berhane *et al.*, "Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 580.
- [15] L. A. V. Ishuaylas, Y. Hirata, L. C. V. Santos, and N. S. Torobeo, "Natural forest mapping in the Andes (Peru): A comparison of the performance of machine-learning algorithms," *Remote Sens.*, vol. 10, no. 5, 2018, Art. no. 782.
- [16] A. Haddadi, M. R. Sahebi, and A. Mansourian, "Polarimetric SAR feature selection using a genetic algorithm," *Can. J. Remote Sens.*, vol. 37, no. 1, pp. 27–36, 2011.
- [17] Y. Zhou, H. Wang, F. Xu, and Y.-Q. Jin, "Polarimetric SAR image classification using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1935–1939, Dec. 2016.
- [18] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 30–42, Jan. 2012.
- [19] L. Zhang, W. Ma, and D. Zhang, "Stacked sparse autoencoder in PolSAR data classification using local spatial information," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1359–1363, Sep. 2016.
- [20] W. Xie *et al.*, "PolSAR image classification via Wishart-AE model or Wishart-CAE model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3604–3615, Aug. 2017.
- [21] Y. Cao, Y. Wu, P. Zhang, W. Liang, and M. Li, "Pixel-wise PolSAR image classification via a novel complex-valued deep fully convolutional network," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2653.
- [22] S. Yang, M. Wang, Z. Feng, Z. Liu, and R. Li, "Deep sparse tensor filtering network for synthetic aperture radar images classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3919–3924, Aug. 2018.
- [23] Z. Wen, Q. Wu, Z. Liu, and Q. Pan, "Polar-spatial feature fusion learning with variational generative-discriminative network for PolSAR classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8914–8927, Nov. 2019.
- [24] B. Zou, X. Xu, and L. Zhang, "Object-based classification of PolSAR images based on spatial and semantic features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 609–619, Jan. 2020.
- [25] H. Liu, Y. Wang, S. Yang, S. Wang, J. Feng, and L. Jiao, "Large polarimetric SAR data semi-supervised classification with spatial-anchor graph," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1439–1458, Apr. 2016.
- [26] W. Fan, F. Zhou, X. Bai, M. Tao, and T. Tian, "Ship detection using deep convolutional neural networks for PolSAR images," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2862.
- [27] G. A. Fricker, J. D. Ventura, J. A. Wolf, M. P. North, F. W. Davis, and J. Franklin, "A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 2326.
- [28] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020.
- [29] H. Dong, L. Zhang, and B. Zou, "PolSAR image classification with lightweight 3D convolutional networks," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 396.
- [30] C. Yang, B. Hou, B. Ren, Y. Hu, and L. Jiao, "CNN-based polarimetric decomposition feature selection for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8796–8812, Nov. 2019.
- [31] A. G. Mullissa, C. Persello, and A. Stein, "PolSARNet: A deep fully convolutional network for polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5300–5309, Dec. 2019.
- [32] W. Xi, U. Jang, L. J. Chen, and S. Jha, "Manifold assumption and defenses against adversarial perturbations," 2017 *arXiv: 1711.08001*.
- [33] S. Uhlmann and S. Kiranyaz, "Integrating color features in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2197–2216, Apr. 2014.
- [34] P. Yu, A. K. Qin, and D. A. Clausi, "Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1302–1317, Apr. 2012.
- [35] J.-S. Lee, M. R. Grunes, and G. de Grandi, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2363–2373, Sep. 1999.



Wenqiang Hua (Member, IEEE) received the B.E. degree in physical electronics from the University of Electronic Science and Technology of China, Chengdu, China, in 2012, and the Ph.D. degree in circuits and systems from the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University, Xi'an, China, in 2018.

Since 2018, he has been with the School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, China, where he is currently a Lecturer. His research interests include machine learning, deep learning, and PolSAR image processing.



Xiaomin Jin received the Ph.D. degree in computer science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2018.

He is currently a Lecturer with the School of Computer Science and Technology, Xi'an University of Post and Telecommunication, Xi'an, China. His research interests include machine learning and deep learning.



Wen Xie (Member, IEEE) received the B.S. degree in automation from Xidian University, Xi'an, China, in 2011, and the Ph.D. degree in circuits and systems from the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University, Xi'an, China, in 2017.

Since 2017, she has been with the School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, China, where she is currently a Lecturer. Her research interests include machine learning, deep learning, and image classification.